Differential Abundance methods- Simulation Results

Saritha Kodikara

February 21, 2022

Contents

1 Simulation Results- Scenario 1

In this file, we present results from the simulations under moderate parameter values (i.e. dispersion=0.1 and AR=0.04) for Zero-inflated Beta Regression Model, Negative Binomial Mixed Model, splinectomeR, Zero-inflated Gaussian mixed model and Fast zero-inflated negative binomial mixed model.

```
meta_df<-readRDS("Data/df_meta.Rdata")
list_sc1_ra<-readRDS("Data/RA_Scenario1.Rdata")
list_sc1_co<-readRDS("Data/count_Scenario1.Rdata")
outputPath1<-"Results/"
outputPath2<-"Results/run_times/"
nSim=50

#True effects
time_effect<-c(rep(1,10),rep(0,10), rep(1,10), rep(0,270))
group_effect<-c(rep(0,10),rep(1,20), rep(0,270))
group_time_effect<-c(rep(0,20),rep(1,10), rep(0,270))
sig=0.05
nworkers <- parallel::detectCores()
BPPARAM <- BiocParallel::MulticoreParam(workers = nworkers)
```

1.1 Zero-inflated Gaussian mixed model

```
#Scenario 1 (Relative Abundance)
zigmm.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  zigmm<-se.sp(list_sc1_ra[[i]],meta_df,
               time_effect,group_effect,group_time_effect,
               correlation_AR=FALSE, countData=FALSE, method="zigmm")
}, BPPARAM = BPPARAM)

zigmm.se_sp_df<-do.call(rbind.data.frame,
                        lapply(zigmm.se_sp, `[[`, 1))
```

```
zigmm.runTime<-do.call(rbind.data.frame,
                       lapply(zigmm.se_sp, `[[`, 2))[,1:3]
colnames(zigmm.runTime)<-c("user",  "system", "elapsed ")
write.csv(zigmm.se_sp_df, file=paste0(outputPath1,"ZIGMM.csv"))
write.csv(zigmm.runTime, file=paste0(outputPath2,"ZIGMM_run.csv"))


#Scenario 1 (Relative Abundance)-with AR
zigmmAr.se_sp <-BiocParallel::bplapply(1:nSim, function(i){
  zigmm<-se.sp(list_sc1_ra[[i]],meta_df,
               time_effect,group_effect,group_time_effect,
               correlation_AR=TRUE, countData=FALSE, method="zigmm")
}, BPPARAM = BPPARAM)

zigmmAr.se_sp_df<-do.call(rbind.data.frame,
                          lapply(zigmmAr.se_sp, `[[`, 1))
zigmmAr.runTime<-do.call(rbind.data.frame,
                         lapply(zigmmAr.se_sp, `[[`, 2))[,1:3]
colnames(zigmmAr.runTime)<-c("user",  "system", "elapsed ")
write.csv(  zigmmAr.se_sp_df, file=paste0(outputPath1,"ZIGMMAr.csv"))
write.csv(zigmmAr.runTime, file=paste0(outputPath2,"ZIGMMAr_run.csv"))

#Scenario 1 (Counts)
zigmmCo.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  zigmm<-se.sp(list_sc1_co[[i]],meta_df,
               time_effect,group_effect,group_time_effect,
               correlation_AR=FALSE, countData=FALSE, method="zigmm")
}, BPPARAM = BPPARAM)

zigmmCo.se_sp_df<-do.call(rbind.data.frame,
                          lapply(zigmmCo.se_sp, `[[`, 1))
zigmmCo.runTime<-do.call(rbind.data.frame,
                         lapply(zigmmCo.se_sp, `[[`, 2))[,1:3]
colnames(zigmmCo.runTime)<-c("user",  "system", "elapsed ")
write.csv(  zigmmCo.se_sp_df, file=paste0(outputPath1,"ZIGMMCo.csv"))
write.csv(zigmmCo.runTime, file=paste0(outputPath2,"ZIGMMCo_run.csv"))

#Scenario 1 (Counts)-with AR
zigmmCoAR.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  zigmm<-se.sp(list_sc1_co[[i]],meta_df,
               time_effect,group_effect,group_time_effect,
               correlation_AR=TRUE, countData=TRUE, method="zigmm")
}, BPPARAM = BPPARAM)

zigmmCoAR.se_sp_df<-do.call(rbind.data.frame,
                            lapply(zigmmCoAR.se_sp, `[[`, 1))
zigmmCoAR.runTime<-do.call(rbind.data.frame,
                           lapply(zigmmCoAR.se_sp, `[[`, 2))[,1:3]
colnames(zigmmCoAR.runTime)<-c("user",  "system", "elapsed ")
write.csv(  zigmmCoAR.se_sp_df, file=paste0(outputPath1,"ZIGMMCoAR.csv"))
write.csv(zigmmCoAR.runTime, file=paste0(outputPath2,"ZIGMMCoAR_run.csv"))
```

1.2 Zero-inflated Beta Regression Model

```
#Scenario 1
zibr.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  zibr<-se.sp_zibr(list_sc1_ra[[i]],meta_df,
                   time_effect,group_effect,group_time_effect)
}, BPPARAM = BPPARAM)

zibr.se_sp_df<-do.call(rbind.data.frame,
                       lapply(zibr.se_sp, `[[`, 1))
zibr.runTime<-do.call(rbind.data.frame,
                      lapply(zibr.se_sp, `[[`, 2))[,1:3]
colnames(zibr.runTime)<-c("user",  "system", "elapsed ")
write.csv(zibr.se_sp_df, file=paste0(outputPath1,"ZIBR.csv"))
write.csv(zibr.runTime, file=paste0(outputPath2,"ZIBR_run.csv"))
```

1.3 Negative Binomial Mixed Model

```
#Scenario 1
nbmm.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  nbmm<-se.sp(list_sc1_co[[i]],meta_df,
              time_effect,group_effect,group_time_effect,
              correlation_AR=FALSE, countData=TRUE, method="nbmm")
}, BPPARAM = BPPARAM)

nbmm.se_sp_df<-do.call(rbind.data.frame,
                       lapply(nbmm.se_sp, `[[`, 1))
nbmm.runTime<-do.call(rbind.data.frame,
                      lapply(nbmm.se_sp, `[[`, 2))[,1:3]
colnames(nbmm.runTime)<-c("user",  "system", "elapsed ")
write.csv(nbmm.se_sp_df, file=paste0(outputPath1,"NBMM.csv"))
write.csv(nbmm.runTime, file=paste0(outputPath2,"NBMM_run.csv"))

#Scenario 1-AR
nbmmAR.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  nbmm<-se.sp(list_sc1_co[[i]],meta_df,
              time_effect,group_effect,group_time_effect,
              correlation_AR=TRUE, countData=TRUE, method="nbmm")
})

nbmmAR.se_sp_df<-do.call(rbind.data.frame,
                         lapply(nbmmAR.se_sp, `[[`, 1))
nbmmAR.runTime<-do.call(rbind.data.frame,
                        lapply(nbmmAR.se_sp, `[[`, 2))[,1:3]
colnames(nbmmAR.runTime)<-c("user",  "system", "elapsed ")
write.csv(nbmmAR.se_sp_df, file=paste0(outputPath1,"NBMMAR.csv"))
write.csv(nbmmAR.runTime, file=paste0(outputPath2,"NBMMAR_run.csv"))
```

1.4 Fast zero-inflated negative binomial mixed model

```
#Scenario 1
zinbmm.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  zinbmm<-se.sp(list_sc1_co[[i]],meta_df,
```

```
                time_effect,group_effect,group_time_effect,
                correlation_AR=FALSE, countData=TRUE, method="zinbmm")
}, BPPARAM = BPPARAM)

zinbmm.se_sp_df<-do.call(rbind.data.frame,
                         lapply(zinbmm.se_sp, `[[`, 1))
zinbmm.runTime<-do.call(rbind.data.frame,
                        lapply(zinbmm.se_sp, `[[`, 2))[,1:3]
colnames(zinbmm.runTime)<-c("user",  "system", "elapsed ")
write.csv(zinbmm.se_sp_df, file=paste0(outputPath1,"ZINBMM.csv"))
write.csv(zinbmm.runTime, file=paste0(outputPath2,"ZINBMM_run.csv"))

#Scenario 1-AR
zinbmmAR.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  zinbmm<-se.sp(list_sc1_co[[i]],meta_df,
                time_effect,group_effect,group_time_effect,
                correlation_AR=TRUE, countData=TRUE, method="zinbmm")
}, BPPARAM = BPPARAM)

zinbmmAR.se_sp_df<-do.call(rbind.data.frame,
                           lapply(zinbmmAR.se_sp, `[[`, 1))
zinbmmAR.runTime<-do.call(rbind.data.frame,
                          lapply(zinbmmAR.se_sp, `[[`, 2))[,1:3]
colnames(zinbmmAR.runTime)<-c("user",  "system", "elapsed ")
write.csv(zinbmmAR.se_sp_df, file=paste0(outputPath1,"ZINBMMAR.csv"))
write.csv(zinbmmAR.runTime, file=paste0(outputPath2,"ZINBMMAR_run.csv"))
```

1.5 SplinectomeR

```
#Scenario 1
SplinectomeR.se_sp <- BiocParallel::bplapply(1:nSim, function(i){
  SplinectomeR<-se.sp_splineR(list_sc1_ra[[i]],meta_df,
                              time_effect,group_effect)
}, BPPARAM = BPPARAM)

SplinectomeR.se_sp_df<-do.call(rbind.data.frame,
                               lapply(SplinectomeR.se_sp, `[[`, 1))
SplinectomeR.runTime<-do.call(rbind.data.frame,
                              lapply(SplinectomeR.se_sp, `[[`, 2))[,1:3]
colnames(SplinectomeR.runTime)<-c("user",  "system", "elapsed ")
write.csv(SplinectomeR.se_sp_df, file=paste0(outputPath1,"SplinectomeR.csv"))
write.csv(SplinectomeR.runTime, file=paste0(outputPath2,"SplinectomeR_run.csv"))
```

```
  setwd("Results/")
  temp_i <-list.files(pattern="*.csv")
  Sc1_boxplot<-boxplotFunction(temp_i, plotTitle = "AR=0.04 & Dispersion=0.1")
  Sc1_boxplot[[1]]+Sc1_boxplot[[2]]+Sc1_boxplot[[3]]+
    plot_layout(guides = "collect") +
  plot_annotation(tag_levels = 'A')
```

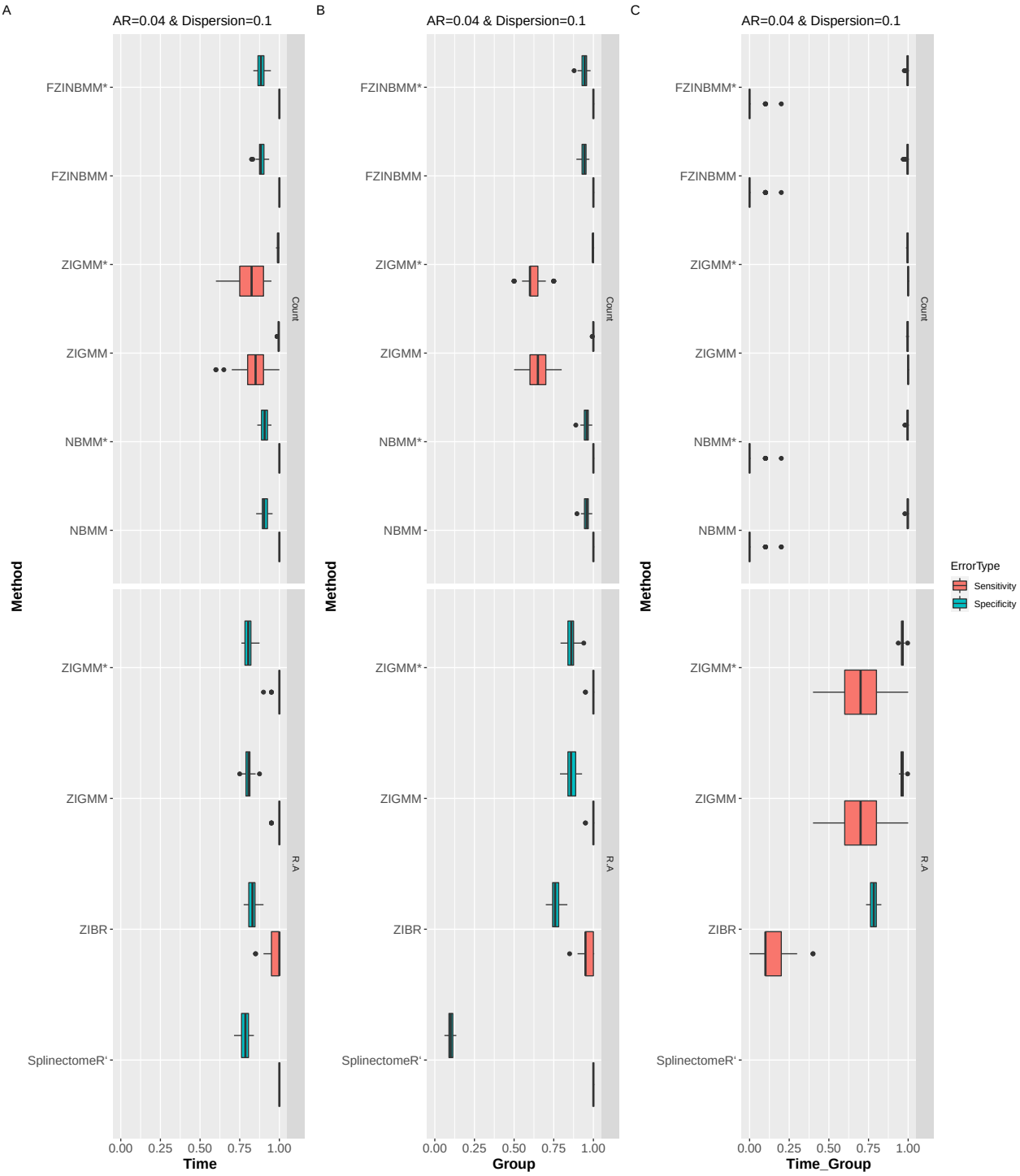

Figure 1: Sensitivity and specificity results from differential abundance methods for time effect (A), group effect (B) and time group interaction effect (C) when AR is o 0.04 and dispersion is 0.1.